

事例研究

清 雄一

株式会社三菱総合研究所

情報技術研究センター

研究員

インシデントデータを活用したプロセス改善に向けて

講演概要

毎月100件を超えるシステム障害（インシデント）が全社で起きている。各インシデントに対して暫定対応や恒久対応は実施しているが、似たようなインシデントが継続的に発生してしまっている。各インシデントへの個別対応のみではなく、開発プロセスや運用プロセスそのものを改善する必要があると考え、プロセス改善のためのプロジェクトを立ち上げた。

今までに各インシデントの発生原因を記載したインシデントデータを6768件収集しており、プロセス改善のためにこのデータを利用できると考えられる。しかし現状はあまり活用されていない。発生原因は自由記述であり、数千件の自由記述のデータを読み込むことは現実的ではないためである。また、仮に各インシデントを逐一読み込み、それぞれに対してプロセス改善策を考えたとしても、全体を見渡した最適な改善を策定することができない。一方、インシデントデータには「表層原因区分」および「深層原因区分」をそれぞれ20項目程度から選択させ記載している。各インシデントがどのような原因で発生したかの大きなカテゴリはわかるが、項目の抽象度が高過ぎ、具体的な対策案は見えてこない。

このような状況を踏まえ、本プロジェクトでは、共通の対策や対策方針を立てることができるほど詳細な原因区分を定義し、各インシデントに対してできるだけ機械的に精度よく付与することを目指している。詳細な原因区分のレベルで同一のインシデントの集合を抽出することにより、複数のインシデントを見渡した、より最適なプロセス改善策を策定できるようになる。また、プロセス改善策は部署ごとに策定するものとする。部署ごとの背景や環境等も考慮する必要があるからである。

まず20件のインシデントを読み込み、40種類程度の原因区分を定義した。原因区分の洗い出しのためには、さらなるデータ解析が必要であるが、人手の作業には限界がある。日本語解析ツールを利用し、インシデントの発生原因を端的によく表すキーワードの抽出技術を開発している。

また、各インシデントに対する原因区分の機械的な付与にあたっては、テキストマイニング技術を利用することにより、高速・高精度に実施することを目指している。新しい原因区分を機械的に付与する際には、現在既に付与されている原因区分を参考にす。しかし、原因区分が付与されているインシデント数は、全体の6768件に対し、わずか1542件のみである。特に部署ごとに見た場合、たとえば原因区分が付与されているインシデント数が50件以上の部署は、全29部署のうち、わずか7部署のみである。テキストマイニング技術を利用することにより、まず現状の原因区分を、未設定の各インシデントに機械的に付与した。その結果、表層原因が付与されているインシデント数が50件以上の部署は7部署から14部署に増加した。また、クロスバリデーションを実施したところ、80%以上の原因区分が一致しており、精度も高いと考えられる。

今後、新規の原因区分の定義を完了させ、各インシデントに機械的に付与していく。単語間の関係等のオントロジーを整備し、また、キーワード抽出手法の改善を行うことにより、マイニングの精度を高めるとともに、インシデントデータの更なる活用方法を模索していく。

F4a

7月27日

14：40～15：25

会議室A